(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 May 2007 (10.05.2007)

PCT

(10) International Publication Number
WO 2007/053719 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C07H 21/04* (2006.01)

(21) International Application Number: PCT/US2006/042739

(22) International Filing Date: 31 October 2006 (31.10.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/732,040 31 October 2005 (31.10.2005) US

(71) Applicant *(for all designated States except US)*: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; West 116th Street and Broadway, New York, NY 10027 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **JU, Jingyue** [US/US]; 267 Marietta St., Englewood Cliffs, NJ 07632 (US). **BI, Lanrong** [US/US]; 610 West 115th St., Apt. 63A, New York, NY 10025 (US). **KIM, Dae, H.** [US/US]; 100 Morningside Dr., Apt. 5H, New York, NY 10027 (US). **MENG, Qingling** [CN/US]; 529 West 111th ST., Apt. 22A, New York, NY 10027 (US).

(74) Agent: **WHITE, John, P.**; COOPER & DUNHAM LLP, 1185 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Applicants: Jingyue Ju
Serial No.: 10/591,520
Filed: March 3, 2005
Exhibit 3

(54) Title: CHEMICALLY CLEAVABLE 3'-O-ALLYL-DNTP-ALLYL-FLUOROPHORE FLUORESCENT NUCLEOTIDE ANALOGUES AND RELATED METHODS

(57) Abstract: This invention provides a nucleotide analogue comprising (i) a base selected from the group consisting of adenine, guanine, cytosine, thymine and uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3'-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker, and methods of nucleic acid sequencing employing the nucleotide analogue.

WO 2007/053719 A2

CHEMICALLY CLEAVABLE 3'-O-ALLYL-DNTP-ALLYL-FLUOROPHORE FLUORESCENT NUCLEOTIDE ANALOGUES AND RELATED METHODS

The invention disclosed herein was made with Government support under Center of Excellence in Genomic Science grant number IP50 HG002806-01 from the National Institutes of Health, U.S. Department of Health and Human Services. Accordingly, the U.S. Government has certain rights in this invention.

Throughout this application, various publications are referenced in parentheses by number. Full citations for these references may be found at the end of each experimental section. The disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains.

Background of the Invention

With the completion of human genome project, there is now a focus on developing new DNA sequencing technology that will reduce the cost of sequencing dramatically without sacrificing accuracy, which will ultimately enable personalized medicine in healthcare (1). Current state-of-the-art DNA sequencing technology faces limitation in terms of cost, read-length, and throughput. In this regard, DNA sequencing by synthesis (SBS), where the identity of each nucleotide is detected immediately after its incorporation into a growing strand of DNA in a

polymerase reaction, offers an alternative approach to address some of these limitations. An important requirement for the SBS approach is a 3'-OH capped fluorescent nucleotide that can act as a reversible terminator (2), where after the identification of the nucleotide incorporated in a DNA polymerase reaction, the 3'-OH capping group along with fluorescent label are removed to regenerate a free 3'-OH group thus allowing DNA chain elongation. The importance of removing the fluorescent label after each base identification is to make sure that the residual fluorescence from the previous nucleotide incorporation does not affect the identification of the next incorporated fluorescent nucleotide.

The speed and sequence read length of SBS depend on the yield of the cleavage efficiency of the fluorophore and the allyl group. Due to multiple steps required in the identification, removal of fluorescent label, and regeneration of 3'-OH group after each nucleotide incorporation in SBS, the loss of even a minor efficiency at each step may lead to inhibition of prolonged read length. For this reason, any improvement in efficiency within each cycle of nucleotide identification, fluorophore removal, and 3'-OH regeneration can have significant impact on read length, thus tackling the physical limits in DNA sequencing by synthesis.

Summary

This invention provides a nucleotide analogue comprising (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine, thymine or an analogue of thymine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3'-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker.

This invention also provides a method for making a nucleotide analogue wherein the nucleotide analogue comprises (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3' oxygen of the deoxyribose, and (iv) a fluorophore bound to the base via an allyl linker which is not an *iso*-allyl linker, comprising the steps of:

(a) contacting 6-amino-hex-2-en-1-ol and an N-hydroxysuccinimide ester of a fluorophore in the presence of a first suitable solvent and a suitable base;

(b) treating the resulting product of step (a) with DSC/Et_3N in a second suitable solvent; and

(c) treating the resulting product of step (b) with a 3'-O-allyl-dNTP-NH_2 in the presence of a suitable buffered solvent, wherein the base of the

3′-O-allyl-dNTP-NH_2 is an adenine, guanine, cytosine, uracil, or an analogue thereof, thereby making the nucleotide analogue.

This invention also provides a method for making a nucleotide analogue wherein the nucleotide analogue comprises (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′ oxygen of the deoxyribose, and (iv) a fluorophore bound to the base via an allyl linker, comprising the steps of:

(a) contacting 2-(2-amino-ethyl)-prop-2-en-1-ol and an N-hydroxysuccinimide ester of a fluorophore in the presence of a first suitable solvent and a suitable base;

(b) treating the resulting product of step (a) with DSC/Et_3N in a second suitable solvent; and

(c) treating the resulting product of step (b) with a 3′-O-allyl-dNTP-NH_2 in the presence of a suitable buffered solvent, wherein the base of the 3′-O-allyl-dNTP-NH_2 is an adenine, guanine, cytosine, uracil, or an analogue thereof, thereby making the nucleotide analogue.

This invention also provides a method for determining the sequence of a DNA comprising performing the following steps for each residue of the DNA to be sequenced:

(a) contacting the DNA with a DNA polymerase in the presence of (i) a primer and (ii) four fluorescent nucleotide analogues under conditions

permitting the DNA polymerase to catalyze DNA synthesis, wherein (1) the nucleotide analogues consist of an analogue of dGTP, an analogue of dCTP, an analogue of dTTP or dUTP, and an analogue of dATP, (2) each nucleotide analogue comprises (i) a base selected from the group consisting of adenine, guanine, cytosine, thymine or uracil, and analogues thereof, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker, so that a nucleotide analogue complementary to the residue being sequenced is bound to the DNA by the DNA polymerase, and (3) each of the four analogues has a predetermined fluorescence wavelength which is different than the fluorescence wavelengths of the other three analogues;

(b) removing unbound nucleotide analogues;

(c) determining the identity of the bound nucleotide analogue; and

(d) following step (c), except with respect to the final DNA residue to be sequenced, chemically cleaving from the bound nucleotide analogue the fluorophore and the allyl moiety bound to the 3′-oxygen atom of the deoxyribose,

thereby determining the sequence of the DNA.

This invention also provides a kit for performing the instant method comprising, in separate compartments,

(a) a nucleotide analogue of (i) GTP, (ii) ATP, (iii) CTP and (iv) TTP or UTP, wherein each analogue comprises (i) a base selected from the

group consisting of adenine, guanine, cytosine, thymine or uracil, or an analogue thereof, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker,

(b) reagents suitable for use in DNA polymerization; and

(c) instructions for use.

This invention also provides a method for covalently affixing a detectable moiety, via an allyl linker, to an NH_2-bearing molecule, comprising contacting the detectable moiety with the the NH_2-bearing molecule in the presence of a suitable solvent and suitable base, wherein the detectable moiety comprises a mass tag, fluorophore or chromophore bound to a NHS ester of an allyl moiety. In one embodiment of the NH_2-bearing molecule is a nucleotide and the detectable moiety comprises a fluorophore.

Brief Description of the Figures

Figure 1: Structures of four 3'-O-allyl-dNTP-PC-fluorophores.

Figure 2. Synthesis of 3'-O-allyl-dNTP-allyl-fluorophores.

Figure 3. Polymerase DNA extension reaction using 3'-O-allyl-dGTP-allyl-Bodipy-FL-510 as a reversible terminator.

Figure 4. A polymerase extension scheme (left) and MALDI-TOF MS spectra of extension and dual-deallylation product (right).

Figure 5. Structures of four alternative chemically cleavable fluorescent nucleotides, 3'-O-allyl-dNTP-iso-allyl-fluorophore.

Figure 6. Synthesis of 3'-O-allyl-dNTP-iso-allyl-fluorophore.

Figure 7. A polymerase extension scheme (left) and MALDI-TOF MS spectra of extension and dual-deallylation product (right).

Detailed Description of the Invention

Terms

The following definitions are presented as an aid in understanding this invention:

A	–	Adenine;
C	–	Cytosine;
DNA	–	Deoxyribonucleic acid;
DMF	–	Dimethylformamide;
G	–	Guanine;
NHS	–	N-hydroxysuccinimidyl;
RNA	–	Ribonucleic acid;
SBS	–	Sequencing by synthesis;
T	–	Thymine; and
U	–	Uracil.

"Nucleic acid" shall mean any nucleic acid, including, without limitation, DNA, RNA and hybrids thereof. The nucleic acid bases that form nucleic acid molecules can be the bases A, C, G, T and U, as well as derivatives thereof. Derivatives of these bases are well known in the art, and are exemplified in PCR Systems, Reagents and Consumables (Perkin Elmer Catalogue 1996 1997, Roche Molecular Systems, Inc., Branchburg, New Jersey, USA).

As used herein, "self-priming moiety" shall mean a nucleic acid moiety covalently bound to a nucleic acid to be transcribed, wherein the bound nucleic acid moiety, through its proximity with the transcription initiation

site of the nucleic acid to be transcribed, permits transcription of the nucleic acid under nucleic acid polymerization-permitting conditions (e.g. the presence of a suitable polymerase, nucleotides and other reagents). That is, the self-priming moiety permits the same result (i.e. transcription) as does a non-bound primer. In one embodiment, the self-priming moiety is a single stranded nucleic acid having a hairpin structure.

"Hybridize" shall mean the annealing of one single-stranded nucleic acid to another nucleic acid based on sequence complementarity. The propensity for hybridization between nucleic acids depends on the temperature and ionic strength of their milieu, the length of the nucleic acids and the degree of complementarity. The effect of these parameters on hybridization is well known in the art (see Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York.)

As used herein, "nucleotide analogue" shall mean an analogue of A, G, C, T or U (that is, an analogue of a nucleotide comprising the base A, G, C, T or U) which is recognized by DNA or RNA polymerase (whichever is applicable) and incorporated into a strand of DNA or RNA (whichever is appropriate). Examples of nucleotide analogues include, without limitation 7-deaza-adenine, 7-deaza-guanine, the analogues of deoxynucleotides shown in Figure 6, analogues in which a label is attached through a cleavable linker to the 5-position of cytosine or thymine or to the 7-position of deaza-adenine or deaza-

guanine, analogues in which a small chemical moiety such as $-CH_2OCH_3$ or $-CH_2CH{=}CH_2$ is used to cap the -OH group at the 3'-position of deoxyribose, and analogues of related dideoxynucleotides. Nucleotide analogues, including dideoxynucleotide analogues, and DNA polymerase-based DNA sequencing are also described in U.S. Patent No. 6,664,079.

1,3 dipolar azide-alkyne cycloaddition chemistry is described in WO 2005/084367 and PCT/US03/39354, the contents of each of which are hereby incorporated by reference.

All embodiments of U.S. Patent No. 6,664,079 (the contents of which are hereby incorporated by reference) with regard to sequencing a nucleic acid are specifically envisioned here.

With regard to the synthesis of the nucleotide analogues disclosed herein, other fluorophores or chromophores to be cleavably attached to the base of the analogue are envisioned. In addition, combinatorial fluorescence energy tags as described in U.S. Patent No. 6,627,748 (the contents of which are hereby incorporated by reference) or mass tags may be used in place of the fluorophores described herein.

Embodiments of the Invention

This invention provides a nucleotide analogue comprising (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine, thymine or an analogue of thymine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker.

In one embodiment, the nucleotide analogue is an analogue of dATP, dGTP, dCTP or dUTP. In one embodiment, the fluorophore is selected from the group consisting of ROX, Bodipy-FL-510, Bodipy-650 and R6G. In one embodiment, the fluorophore is bound to the base via an *iso*-allyl linker.

In a further embodiment, the nucleotide analogue is selected from the group consisting of 3′-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-*iso*-allyl-Bodipy-650, 3′-O-allyl-dATP-*iso*-allyl-ROX and 3′-O-allyl-dUTP-*iso*-allyl-R6G, or is selected from the group consisting of 3′-O-allyl-dGTP-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-allyl-Bodipy-650, 3′-O-allyl-dATP-allyl-ROX and 3′-O-allyl-dUTP-allyl-R6G.

This invention also provides a method for making a nucleotide analogue wherein the nucleotide analogue comprises (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′ oxygen of the deoxyribose, and (iv) a fluorophore bound to the base via an allyl linker which is not an *iso*-allyl linker, comprising the steps of:

(a) contacting 6-amino-hex-2-en-1-ol and an N-hydroxysuccinimide ester of a fluorophore in the presence of a first suitable solvent and a suitable base;

(b) treating the resulting product of step (a) with DSC/Et_3N in a second suitable solvent; and

(c) treating the resulting product of step (b) with a 3′-O-allyl-dNTP-NH_2 in the presence of a suitable buffered solvent, wherein the base of the 3′-O-allyl-dNTP-NH_2 is an adenine, guanine, cytosine, uracil, or an analogue thereof, thereby making the nucleotide analogue.

In one embodiment, the steps of the instant method comprise those set forth in Figure 2, scheme A; Figure 2, scheme B; Figure 2, scheme C; or Figure 2, scheme D.

In one embodiment of the instant method, the first suitable solvent is DMF and the second suitable solvent is DMF, and in another embodiment the first suitable solvent is acetonitrile and the second suitable solvent

is DMF. In one embodiment of the instant method, the suitable base is $NaHCO_3$. In one embodiment of the instant method, the suitable buffered solvent is DMF buffered with $NaHCO_3$-Na_2CO_3. In another embodiment of the instant method, the suitable buffered solvent is acetonitrile buffered with $NaHCO_3$-Na_2CO_3.

This invention also provides a method for making a nucleotide analogue wherein the nucleotide analogue comprises (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′ oxygen of the deoxyribose, and (iv) a fluorophore bound to the base via an allyl linker, comprising the steps of:

(a) contacting 2-(2-amino-ethyl)-prop-2-en-1-ol and an N-hydroxysuccinimide ester of a fluorophore in the presence of a first suitable solvent and a suitable base;

(b) treating the resulting product of step (a) with DSC/Et_3N in a second suitable solvent; and

(c) treating the resulting product of step (b) with a 3′-O-allyl-dNTP-NH_2 in the presence of a suitable buffered solvent, wherein the base of the 3′-O-allyl-dNTP-NH_2 is an adenine, guanine, cytosine, uracil, or an analogue thereof, thereby making the nucleotide analogue.

In one embodiment of the instant method, the steps comprise those set forth in Figure 6, scheme A; Figure 6, scheme B; Figure 6, scheme C; or Figure 6, scheme D.

In one embodiment of the instant method, the first suitable solvent is DMF and the second suitable solvent is DMF. In one embodiment of the instant method, the suitable base is $NaHCO_3$. In one embodiment of the instant method, the suitable buffered solvent is DMF buffered with $NaHCO_3$-Na_2CO_3.

This invention also provides a method for determining the sequence of a DNA comprising performing the following steps for each residue of the DNA to be sequenced:

(a) contacting the DNA with a DNA polymerase in the presence of (i) a primer and (ii) four fluorescent nucleotide analogues under conditions permitting the DNA polymerase to catalyze DNA synthesis, wherein (1) the nucleotide analogues consist of an analogue of dGTP, an analogue of dCTP, an analogue of dTTP or dUTP, and an analogue of dATP, (2) each nucleotide analogue comprises (i) a base selected from the group consisting of adenine, guanine, cytosine, thymine or uracil, and analogues thereof, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3'-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker, so that a nucleotide analogue complementary to the residue being sequenced is bound to the DNA by the DNA polymerase, and (3) each of the four analogues has a predetermined fluorescence wavelength which is different than the fluorescence wavelengths of the other three analogues;

(b) removing unbound nucleotide analogues;

(c) determining the identity of the bound nucleotide analogue; and

(d) following step (c), except with respect to the final DNA residue to be sequenced, chemically cleaving from the bound nucleotide analogue the fluorophore and the allyl moiety bound to the 3′-oxygen atom of the deoxyribose,

thereby determining the sequence of the DNA.

In one embodiment of the instant method, chemically cleaving the fluorophore and the allyl moiety bound to the 3′-oxygen atom is performed using Na_2PdCl_4.

In one embodiment of the instant method, the primer is a self-priming moiety.

In one embodiment of the instant method, the DNA is bound to a solid substrate. In one embodiment of the instant method, the DNA is bound to the solid substrate via 1,3-dipolar azide-alkyne cycloaddition chemistry. In one embodiment of the instant method, about 1000 or fewer copies of the DNA are bound to the solid substrate.

In one embodiment of the instant method, the four fluorescent nucleotide analogues are 3′-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-*iso*-allyl-Bodipy-650, 3′-O-allyl-dATP-*iso*-allyl-ROX and 3′-O-allyl-dUTP-*iso*-allyl-R6G.

In another embodiment of the instant method, the four fluorescent nucleotide analogues are 3′-O-allyl-dGTP-

allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-allyl-Bodipy-650, 3′-O-allyl-dATP-allyl-ROX and 3′-O-allyl-dUTP-allyl-R6G.

In one embodiment of the instant method, the DNA polymerase is a 9°N polymerase.

This invention also provides a kit for performing the instant method comprising, in separate compartments,

(a) a nucleotide analogue of (i) GTP, (ii) ATP, (iii) CTP and (iv) TTP or UTP, wherein each analogue comprises (i) a base selected from the group consisting of adenine, guanine, cytosine, thymine or uracil, or an analogue thereof, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker,

(b) reagents suitable for use in DNA polymerization; and

(c) instructions for use.

In one embodiment, the kit comprises 3′-O-allyl-dGTP-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-allyl-Bodipy-650, 3′-O-allyl-dATP-allyl-ROX and 3′-O-allyl-dUTP-allyl-R6G.

In one embodiment, the kit comprises 3′-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-*iso*-allyl-Bodipy-650, 3′-O-allyl-dATP-*iso*-allyl-ROX and 3′-O-allyl-dUTP-*iso*-allyl-R6G.

This invention also provides a method for covalently affixing a detectable moiety, via an allyl linker, to an NH_2-bearing molecule, comprising contacting the

detectable moiety with the the NH_2-bearing molecule in the presence of a suitable solvent and suitable base, wherein the detectable moiety comprises a mass tag, fluorophore or chromophore bound to a NHS ester of an allyl moiety. In one embodiment, the NH_2-bearing molecule is a nucleotide and the detectable moiety comprises a fluorophore.

In an embodiment, the allyl is chemically cleaved using a palladium catalyst. In an embodiment the cleaving is performed using Na_2PdCl_4 and TPPTS. In one embodiment the pH is between 8.5 and 9. In a further embodiment the pH is 8.8.

In embodiments of this invention, the sequencing methods described can be applied, mutatis mutandis, to sequencing an RNA molecule or an RNA/DNA hybrid molecule.

This invention will be better understood by reference to the Experimental Details which follow, but those skilled in the art will readily appreciate that the specific experiments detailed are only illustrative of the invention as described more fully in the claims which follow thereafter.

Experimental Details

Synopsis

Here, the construction of a novel chemically cleavable fluorescent labeling system based on an allyl group to modify nucleotides for DNA sequencing by synthesis (SBS) is explored. It is found that an allyl moiety can be used successfully as a linker to tether a fluorophore to a 3'-O-allyl-modified nucleotides (A, C, G, U), forming chemical cleavable reversible terminators, 3'-O-allyl-dNTP-allyl-fluorophore, for application in SBS. The fluorophore and the 3'-O-allyl group on a DNA extension product, which is generated by incorporating 3'-O-allyl-dNTP-allyl-fluorophore, are removed simultaneously in 30 sec by Pd-catalyzed deallylation in aqueous buffer solution. This one-pot dual-deallylation reaction thus allows the re-initiation of the polymerase reaction and increases the SBS efficiency. Expansion of this novel linker and selective protection strategy to other applications that include bio-conjugation, solution- and solid-phase organic synthesis is envisaged.

Introduction

A disulfide group has been previously explored as a chemically cleavable linker to attach a fluorophore to a deoxynucleotide and the use of 2-mercaptoethanol to remove the fluorophore after the nucleotide incorporation and detection in SBS (4). However, the disulfide bond can be reversed and destabilized under certain conditions (5,6). Here, the construction of a novel chemically

cleavable fluorescent labeling system based on an allyl group is disclosed. The discovery permits fluorophore linker cleaving the 3′-O-allyl capping group removal in a single step, thus increasing SBS efficiency. Disclosed here is an allyl moiety that can be used successfully as a linker to tether a fluorophore to a 3′-O-allyl-capped nucleotide, thus forming a set of chemical cleavable reversible terminators, 3′-O-allyl-dNTP-allyl-fluorophores (Fig. 2).

The fluorophore and the 3′-O-allyl group on a DNA extension product which is generated by incorporating structure **4** (Fig. 2) are removed simultaneously in 30 sec by Pd-catalyzed deallylation in aqueous buffer solution. This one-pot dual-deallylation reaction thus allows the re-initiation of the polymerase reaction. Design, synthesis and evaluation of a 3′-O-allyl fluorescent nucleotide, 3′-O-allyl-dGTP-allyl-Bodipy-FL-510 (structure **4**) with a fluorophore attached to the 7 position of guanine base via an allyl carbamate linker, and its application as a reversible terminator for SBS is shown in Fig. 3.

Readily available allylic alcohol (structure **1**) was chosen as a starting material for the preparation of **4**. First, allylic alcohol **1** was reacted with N-hydroxysuccinimide (NHS) ester of the BODIPY-FL-510 to produce allylic-Bodipy-FL-510-NHS (structure **2**), which was subsequently converted to its corresponding NHS ester (structure **3**) by reacting with N,N′-disuccinimidyl carbonate. The coupling reaction between **3** and the modified nucleotide (3′-O-allyl-dGTP-NH_2) (3) produced

the chemically cleavable fluorescent nucleotide, 3'-O-allyl-dGTP-allyl-Bodipy-FL-510 **4**.

To verify that 3'-O-allyl-dGTP-allyl-Bodipy-FL-510 **4**, acting as a reversible terminator, is incorporated accurately in a base-specific manner in a polymerase reaction, a polymerase DNA extension reaction was performed in solution as shown in Fig. 3. This allows the isolation of the DNA product at each step for detailed molecular structure characterization by using MALDI-TOF MS as shown in Fig. 4. First, a polymerase extension reaction using 3'-O-allyl-dGTP-allyl-Bodipy-FL-510 **4** as a terminator along with a primer and the synthetic 100-mer DNA template corresponding to a portion of exon 7 of the human *p53* gene was performed producing a single-base extension product (structure **5**). After the reaction, a small portion of the DNA extension product **5** was characterized by MALDI-TOF MS. The rest of the extended DNA product **5** was added to a deallylation cocktail [1X Thermopol reaction buffer/Na_2PdCl_4/$P(PhSO_3Na)_3$] to perform dual-deallylation in a one-pot reaction for 30 sec to yield deallylated DNA product **6** and characterized by MALDI-TOF MS. The deallylated DNA product with both the fluorophore removed and a free 3'-OH group regenerated can then be used as a primer for a next nucleotide extension reaction.

Fig. 4 (right panel) shows sequential mass spectrum at each step of DNA sequencing by synthesis using 3'-O-allyl-dGTP-allyl-Bodipy-FL-510 **4** as a reversible terminator. As can be seen from Fig. 4A, the MALDI-TOF MS spectrum consists of a distinct peak at *m/z* 6,967

corresponding to the single base DNA extension product **5** with 100% incorporation efficiency, confirming that the reversible terminator **4** can be incorporated base-specifically by DNA polymerase into a growing DNA strand. Fig. 4B shows the one-pot dual-deallylation result after 30 sec incubation of the DNA extension product in a deallylation cocktail solution. The peak at *m/z* 6,967 has completely disappeared, whereas the peak at *m/z* 6,512 corresponding to a DNA product **6** with both the fluorophore and 3′-O-allyl removed appears as the sole product. The absence of a peak at *m/z* 6,967 proves that the one-pot dual-deallylation reaction in removing both the fluorophore and the 3′-O-allyl group from the DNA product was completed with high efficiency. Furthermore, the allyl linker moiety is completely stable in a polymerase extension condition and can be selectively cleaved in a rapid and efficient manner.

Another version of the allyl linker for the construction of the chemically cleavable fluorescent nucleotides as reversible terminators was investigated. The structures of the four molecules (3′-O-allyl-dNTP-*iso*-allyl-fluorophore) are shown in Fig. 5. The syntheses of these molecules are shown in Fig. 6. As an example, the polymerase extension scheme using 3′-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510 as a reversible terminator and the corresponding MALDI-TOF MS results are shown in Fig. 7. The other three chemically fluorescent nucleotides (A, C, U) are similarly characterizable by MALDI-TOF MS.

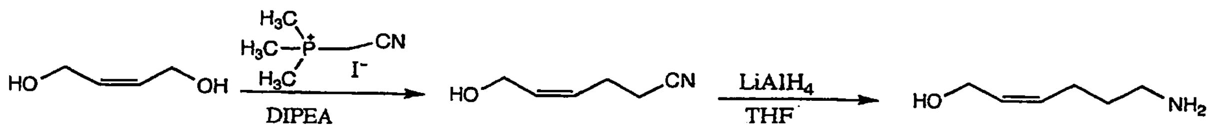
The experimental results show that the 3′-O-allyl-dNTP-allyl-fluorophore can be faithfully incorporated into a

growing DNA strand in a polymerase extension reaction to act as reversible terminators in SBS. This novel linker and selective protection strategy can be applied to other tasks that include bio-conjugation, solution- and solid-phase organic synthesis. .

Material and Methods

General Information. ^{1}H NMR spectra were recorded on a Brucker DPX-400 (400 MHz) spectrometer and are reported in ppm from CD_3OD or $CDCl_3$ internal standard (3.31 or 7.26 ppm respectively). Data are reported as follows: (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, coupling constant(s) in Hz; integration; assignment). Proton decoupled ^{13}C NMR spectra were recorded on a Brucker DPX-400 (100 MHz) spectrometer and are reported in ppm from CD_3OD or $CDCl_3$ internal standard (49.0 or 77.0 ppm respectively). High Resolution Mass Spectra (HRMS) were obtained on a JEOL JMS HX 110A mass spectrometer. Mass measurement of DNA was made on a Voyager DE MALDI-TOF mass spectrometer (Applied Biosystems). The NHS esters of the fluorophores were purchased from Molecular Probes. All other chemicals were purchased from Sigma-Aldrich. 9°N polymerase (exo-) A485L/Y409V was generously provided by New England Biolabs.

Synthesis of chemically cleavable linker: 6-Amino-hexen-1-ol (1)



1

To a mixture of (cis)-2-butene-1,4-diol (440 mg, 5.00 mmol) and (cyanomethyl)-trim thylphosphonium iodide (1.24 g, 5.10 mmol) (Zaragoza, F. J. Org. Chem. 2002; 67(14), 4963-4964) were added propionitrile (4.0 mL) and N,N'-diisopropylethylamine (DIPEA) (1.10 mL, 6.32 mmol), and the mixture was stirred at 97 °C for 24 h. Water (0.20 mL, 11. 1 mmol) was added, and stirring at 97 °C was continued for 15 h. Water (25 mL) and concentrated hydrochloric acid (1.0 mL, 12 mmol) were added, and the mixture was extracted with ethyl acetate (3 X 25 mL). The combined extracts were washed with brine, dried with Na_2SO_4, and concentrated. The residue was purified by flash chromatography to yield 410 mg (74 %) 6-hydroxy-4-hexenenitrile as an oil. ^1H NMR (400 MHz , $CDCl_3$) δ 5.71(m, 1H), 5.62 (m, 1H), 4.20 (d, 2H), 2.45 (t, 2H), 2.33 (q, 2H). ^{13}C NMR (100 MHz, $CDCl_3$) d 129.3, 129.1, 117.7, 60.0, 22.5, 18.3. HRMS m/z: calcd for C_6H_9NO ($M+H^+$) 112.068, found 112.082.

To a suspension of $LiAlH_4$ (380 mg, 10.0 mmol) in THF (50 mL) was added slowly dropwise a solution of 6-hydroxy-4-hexenenitrile (333 mg, 3.00 mmol) in THF (20 mL) while keeping the temperature below 0 °C. When the reaction slowed down, the mixture was heated to reflux for 24 h. The excess $LiAlH_4$ was quenched by addition of 15% sodium hydroxide. The resulting white precipitate was filtered. The filtrate was dried over Na_2SO_4 and concentrated under

reduced pressure to yield 6-Amino-hexen-1-ol **1** (296 mg, 85%) as yellow oil. ^1H NMR (400 MHz, $CDCl_3$) δ 1.54 (q, 2H), 1.82 (brs, 3H, NH_2+OH), 5.71 (m, 1H), 5.62 (m, 1H), 4.20 (d, 2H), 4.07 (d, 2H), 2.70 (t, 2H), 2.09 (q, 2H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 132.3, 129.8, 63.6, 41.8, 32.2, 29.8. HRMS *m/z* calcd for $C_6H_{14}NO$ (M+ H^+) 117.075, found 117.107.

Allylic-Bodipy-FL-510 (2). Allylic alcohol **(1)** (3 mg, 0.026 mmol) was dissolved in 550 µl of acetonitrile and 100 µl of 1 M $NaHCO_3$ aqueous solution. A solution of Bodipy-FL-510 N-hydroxysuccinimidyl (NHS) ester (5 mg, 0.013 mmol) in 400 µl of acetonitrile was added slowly to the above reaction mixture and then stirred for 5 h at room temperature. The resulting reaction mixture was purified on a preparative silica-gel TLC plate ($CHCl_3/CH_3OH$ = 95:5) to give pure allylic-Bodipy-FL-510 **(2)** (4.8 mg, 46%). ^1H NMR (400 MHz, CD_3OD) δ 7.50 (s, 1H), 6.63 (s, 1H), 6.33 (s, 1H), 5.68(m, 1H), 5.62 (m, 1H), 5.45 (m, 2H), 4.21 (d, 2H), 3.23 (t, 2H), 2.24 (t, 2H), 2.22 (t, 2H), 2.16(d, 1H), 2.05 (d, 1H), 2.00 (m, 2H), 1.57 (m, 2H). ^{13}C NMR (100 MHz, CD_3OD) δ 175.4, 163.7, 146.9, 145.8, 132.1, 130.0, 129.2, 128.7, 123.2, 116.0, 109.2, 101.0, 59.4, 40.4, 34.0, 30.6, 30.3, 24.8, 18.7, 16.9. High-resolution MS (FAB^+) *m/z*: anal. calcd for $C_{20}H_{26}O_2N_3F_2B$ (M + H^+), 390.2086; found, 390.2101.

Allylic-Bodipy-FL-510 NHS ester (3). N,N'-disuccinimidyl carbonate (4.27 mg, 0.017 mmol) and triethylamine (4.6 µl, 0.033 mmol) were added to a solution of allylic-Bodipy-FL-510 **(2)** (4.8 mg, 0.012 mmol) in 200 µl of dry

acetonitrile. The reaction mixture was stirred under argon at room temperature for 6 h. Solvent was removed under vacuum, and 1 ml of 1 M $NaHCO_3$ aqueous solution was added to the residual mixture. Extracted with ethyl acetate three times, the combined organic layer was dried over Na_2SO_4, which was directly subjected to the following coupling reaction without further purification.

3'-O-allyl-dGTP-allyl-Bodipy-FL-510 (4). Crude allylic-Bodipy-FL-510 NHS ester **(3)** (6.3 mg) in 300 µl of acetonitrile was added to a solution of 3'-O-allyl-dGTP-NH_2 (2 mg, 0.004 mmol) in 300 µl of Na_2CO_3/$NaHCO_3$ buffer (0.1 M, pH 8.7). The reaction mixture was stirred at room temperature for 3 h, which was subsequently purified by preparative silica-gel TLC plate ($CHCl_3$/CH_3OH, 1/1) to remove unreacted allylic-Bodipy-FL-510 NHS ester **(3)**. The crude product was concentrated further under vacuum and purified with reverse-phase HPLC on a 150 x 4.6-mm C18 column to obtain the pure product 3'-O-allyl-dGTP-allyl-Bodipy-FL-510 **(4)** (retention time of 35 min). Mobile phase: A, 8.6 mM triethylamine/100 mM hexafluoroisopropyl alcohol in water (pH 8.1); B, methanol. Elution was performed with 100% A isocratic over 10 min, followed by a linear gradient of 0-50% B for 20 min and then 50% B isocratic over another 20 min. **4** was characterized by the following primer extension reaction and MALDI-TOF MS.

Primer extension using 3'-O-allyl-dGTP-allyl-Bodipy-FL-510 (4). The polymerase extension reaction mixture consisted of 60 pmol of primer (5'-GTTGAT-GTACACATTGTCAA-3') (SEQ ID NO:1), 80 pmol of 100-mer template (5'-TACCCGGAGGC-

CAAGTACGGCGGGTACGTCCTTGACAATGTGTACATCAACATCACCTACCACCATGTCAGTCTCGGTTGGATCCTCTATTGTGTCCGGG-3') (SEQ ID NO:2), 120 pmol of 3'-O-allyl-dGTP-allyl-Bodipy-FL-510 (**4**), 1X Thermopol reaction buffer (20 mM Tris-HCl/10 mM $(NH_4)_2SO_4$/10 mM KCl/2 mM $MgSO_4$/0.1 % Triton X-100, pH 8.8, New England Biolabs), and 6 units of 9°N Polymerase (exo-)A485L/Y409V in a total volume of 20 µl. The reaction consisted of 20 cycles at 94 °C for 20 sec, 46 °C for 40 sec, and 60 °C for 90 sec. After the reaction, a small portion of the DNA extension product was desalted by using ZipTip and analyzed by MALDI-TOF MS, which shows a dominant peak at *m/z* 6,967 corresponding to the DNA product (**5**). The rest of the product **5** was subjected to the following deallylation.

One-pot dual-deallylation of DNA extension product (5) to produce DNA product (6). DNA product **5** (20 pmol) was added to a mixture of degassed 1X Thermopol reaction buffer (20 mM Tris-HCl/10 mM $(NH_4)_2SO_4$/10 mM KCl/2 mM $MgSO_4$/0.1 % Triton X-100, pH 8.8, 1 µl), Na_2PdCl_4 in degassed H_2O (7 µl, 23 nmol) and $P(PhSO_3Na)_3$ in degassed H_2O (10 µl, 176 nmol) to perform a one-pot dual-deallylation reaction. The reaction mixture was then placed in a heating block and incubated at 70 °C for 30 seconds to yield quantitatively deallylated DNA product (**6**) and analyzed by MALDI-TOF MS to yield a single peak at *m/z* 6,512.

3'-O-allyl-dUTP-allyl-R6G (7): Crude Allylic-R6G NHS ester (prepared by the same procedure as Allylic-Bodipy-FL-510 NHS ester) (7 mg) in 300 µl of DMF was added to a

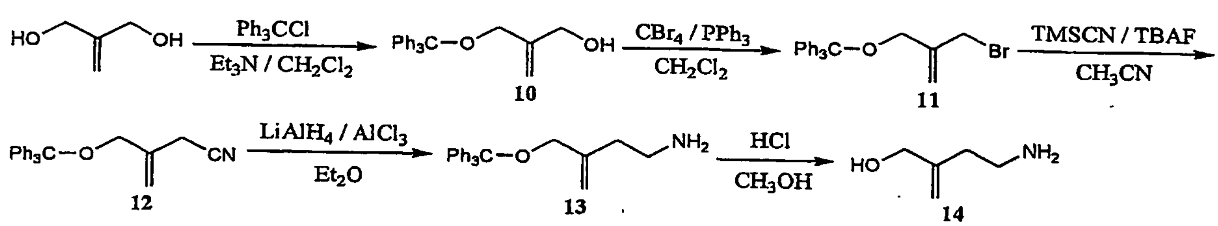
solution of 3'-O-allyl-dUTP-NH_2 (2 mg, 4 μmol) in 300 μl of Na_2CO_3-$NaHCO_3$ buffer (0.1 M, pH 8.7). The reaction mixture was stirred at room temperature for 3 h. A preparative silica-gel TLC plate was used to purify the crude product ($CHCl_3$/CH_3OH, 1/1). The resulting product was concentrated under vacuum and further purified with reverse-phase HPLC on a 150 x 4.6-mm C18 column to obtain the pure product 10 (retention time of 38 min). Mobile phase: A, 8.6 mM triethylamine/100 mM hexafluoroisopropyl alcohol in water (pH 8.1); B, methanol. Elution was performed with 100% A isocratic over 10 min, followed by a linear gradient of 0-50% B for 20 min and then 50% B isocratic over another 20 min. 3'-O-allyl-dUTP-allyl-R6G was characterized by single-base extension reaction and MALDI-TOF MS similarly as for **4**.

3′-O-allyl-dATP-allyl-ROX (8): Crude Allylic-ROX NHS ester (prepared by the same procedure as Allylic-Bodipy-FL-510 NHS ester) (7 mg) in 300 μl of DMF was added to a solution of 3'-O-allyl-dATP-NH_2 (2 mg, 4 μmol) in 300 μl of Na_2CO_3-$NaHCO_3$ buffer (0.1 M, pH 8.7). The reaction mixture was stirred at room temperature for 3 h. A preparative silica-gel TLC plate was used to purify the crude product ($CHCl_3$/CH_3OH, 1/1). The resulting product was concentrated under vacuum and further purified with reverse-phase HPLC on a 150 x 4.6-mm C18 column to obtain the pure product 10 (retention time of 40 min). Mobile phase: A, 8.6 mM triethylamine/100 mM hexafluoroisopropyl alcohol in water (pH 8.1); B, methanol. Elution was performed with 100% A isocratic over 10 min, followed by a linear gradient of 0-50% B for 20 min and then 50% B isocratic over another 20 min. 3'-O-allyl-dATP-allyl-ROX

was characterized by single-base extension reaction and MALDI-TOF MS similarly as for **4**.

3'-O-allyl-dCTP-allyl-Bodipy-650 (9): Crude Allylic-Bodipy-650 NHS ester (prepared by the same procedure as Allylic-Bodipy-FL-510 NHS ester) (7 mg) in 300 µl of DMF was added to a solution of 3'-O-allyl-dCTP-NH_2 (2 mg, 4 µmol) in 300 µl of Na_2CO_3-$NaHCO_3$ buffer (0.1 M, pH 8.7). The reaction mixture was stirred at room temperature for 3 h. A preparative silica-gel TLC plate was used to purify the crude product ($CHCl_3$/CH_3OH, 1/1). The resulting product was concentrated under vacuum and further purified with reverse-phase HPLC on a 150 x 4.6-mm C18 column to obtain the pure product 10 (retention time of 35 min). Mobile phase: A, 8.6 mM triethylamine/100 mM hexafluoroisopropyl alcohol in water (pH 8.1); B, methanol. Elution was performed with 100% A isocratic over 10 min, followed by a linear gradient of 0-50% B for 20 min and then 50% B isocratic over another 20 min. 3'-O-allyl-dCTP-allyl-Bodipy-650 was characterized by single-base extension reaction and MALDI-TOF MS similarly as for **4**.

Synthesis of chemically cleavable linker 4-amino-2-methylene-1-butanol (14):



2-triphenylmethoxymethyl-2-propen-1-ol (10): To a solution of trityl chloride (4.05 g; 14.3 mmol) and 2-methylenepropane-1,3-diol (1.20 mL; 14.3 mmol) in dry CH_2Cl_2 (20 mL) was added triethylamine (4.0 mL; 28.5 mmol) slowly at room temperature. The reaction was stirred at room temperature for 1 h and then ethyl acetate (100 mL) and saturated aqueous $NaHCO_3$ (30 mL) were added. The organic layer was separated and washed by saturated aqueous $NaHCO_3$ and NaCl respectively, and dried over anhydrous Na_2SO_4. After evaporation of the solvent, the residue was purified by flash column chromatography over silica gel using ethyl acetate-hexane (1:10~5) as the eluent to afford **10** as white solid (2.89 g; 62% yield): ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42-7.48 (m, 6H, six of ArH), 7.27-7.33 (m, 6H, six of ArH), 7.20-7.27 (m, 3H, three of ArH), 5.26 (s, 1H, one of $C{=}CH_2$), 5.17 (s, 1H, one of $C{=}CH_2$), 4.13 (d, J = 6.1 Hz, 2H, CH_2OH), 3.70 (s, 2H, Ph_3COCH_2), 1.72 (t, J = 6.1 Hz, 1H, CH_2OH); ^{13}C NMR (100 MHz, $CDCl_3$) δ 145.4, 143.6, 128.3, 127.6, 126.8, 111.6, 87.0, 65.3, 64.5, (8).

1-bromo-2-triphenylmethoxymethyl-2-propene (11): To a solution of **10** (2.56 g; 7.74 mmol) in CH_2Cl_2 (75 ml), CBr_4 (3.63 g; 10.83 mmol) and triphenylphosphine (2.47 g; 9.31 mmol) were added respectively at 0 °C and the reaction was stirred at room temperature for 40 min. Cooled to 0 °C, ethyl acetate (100 mL) and saturated aqueous $NaHCO_3$ (30 mL) were added. The organic layer was separated and washed by saturated aqueous NaCl, and dried over anhydrous Na_2SO_4. After evaporation of the solvent, the residue was purified by flash column chromatography over

silica gel using CH_2Cl_2-hexane (1:5) as the eluent to afford **11** as white solid (3.02 g; 92% yield): ^1H NMR (400 MHz, $CDCl_3$) δ 7.42-7.48 (m, 6H, six of ArH), 7.27-7.33 (m, 6H, six of ArH), 7.20-7.27 (m, 3H, three of ArH), 5.37 (s, 1H, one of C=CH_2), 5.31 (s, 1H, one of C=CH_2), 4.01 (s, 2H, CH_2Br), 3.75 (s, 2H, Ph_3COCH_2); ^{13}C NMR (100 MHz, $CDCl_3$) δ 143.6, 142.6, 128.4, 127.6, 126.9, 115.8, 86.9, 64.2, 33.5, (9).

3-triphenylmethoxymethyl-3-butenenitrile (12): To a solution of **11** (1.45 g; 3.69 mmol) and in dry CH_3CN (37 mL) was added trimethylsilyl cyanide (0.49 mL; 3.69 mmol). Then 1 M tetrabutylammonium fluoride (TBAF) in THF solution (3.69 mL, 3.69 mmol) was added slowly at room temperature and the reaction was stirred for 20 min. Most solvents were evaporated and ethyl acetate (100 mL) and saturated aqueous $NaHCO_3$ (30 mL) were added. The organic layer was washed by saturated aqueous NaCl and dried over anhydrous Na_2SO_4. After evaporation of the solvent, the residue was purified by flash column chromatography over silica gel using acetate-hexane (1:10) as the eluent to afford **12** as white solid (1.01 g; 64% yield): ^1H NMR (400 MHz, $CDCl_3$) δ 7.39-7.45 (m, 6H, six of ArH), 7.21-7.34 (m, 9H, nine of ArH), 5.31 (s, 2H, C=CH_2), 3.64 (s, 2H, Ph_3COCH_2), 3.11 (s, 2H, CH_2CN); ^{13}C NMR (100 MHz, $CDCl_3$) δ 143.3, 135.5, 128.2, 127.7, 126.9, 116.8, 114.7, 87.0, 65.7, 21.9, (10).

3-triphenylmethoxymethyl-3-buten-1-amine (13): To a solution of $LiAlH_4$ (119 mg; 2.98 mmol) in dry ether (5 mL), $AlCl_3$ (400 mg; 2.94 mmol) was added carefully at 0 °C

and the mixture was stirred for 15 min. Then a solution of **12** (829 mg; 2.44 mmol) in dry ether (9 mL) was added and the reaction was stirred at 0 °C for 3h. After that 10% aqueous NaOH (10 mL) was added to quench the reaction. The organic layer was separated and washed by saturated aqueous $NaHCO_3$ and NaCl respectively, and dried over anhydrous K_2CO_3. After evaporation of the solvent, the residue was purified by flash column chromatography over silica gel using CH_3OH-CH_2Cl_2 (1:20~5) as the eluent to afford **10** as colorless oil (545 mg; 65% yield): ^1H NMR (400 MHz, $CDCl_3$) δ 7.41-7.48 (m, 6H, six of ArH), 7.26-7.33 (m, 6H, six of ArH), 7.19-7.26 (m, 3H, three of ArH), 5.33 (s, 1H, one of C=CH_2), 4.96 (s, 1H, one of C=CH_2), 3.53 (s, 2H, Ph_3COCH_2), 2.70 (m, 2H, $CH_2CH_2NH_2$), 2.18 (t, J = 6.7 Hz, 2H, $CH_2CH_2NH_2$), 2.06 (br s, 2H, NH_2); ^{13}C NMR (100 MHz, $CDCl_3$) δ 143.6, 143.4, 128.1, 127.4, 126.5, 111.3, 86.5, 66.1, 39.8, 37.4; HRMS (FAB+) calcd for $C_{24}H_{26}ON$ ($M+H^+$): 344.2014, found: 344.2017, (11).

4-amino-2-methylene-1-butanol (14): To a solution of **13** (540 mg; 1.57 mmol) in CH_3OH (11 mL) was added HCl (2M solution in ether; 5.5 mL) at room temperature and the reaction was stirred for 1 h. Then 7 M methanolic ammonia solution (2.7 mL) was added at room temperature and the reaction was stirred for 10 min. The white solid formed was removed by filtration and washed by CH_3OH. The combined solution was evaporated the residue was purified by flash column chromatography over silica gel using CH_3OH-CH_2Cl_2 (1:4) as the eluent to afford **14** as colorless oil (151 mg; 95% yield): ^1H NMR (400 MHz, CD_3OD) δ 5.19 (s, 1H, one of C=CH_2), 5.01 (s, 1H, one of C=CH_2), 4.06

(s, 2H, C*H*$_2$OH), 3.10 (t, *J* = 7.5 Hz, 2H, CH$_2$C*H*$_2$NH$_2$), 2.46 (t, *J* = 7.5 Hz, 2H, C*H*$_2$CH$_2$NH$_2$); ^{13}C NMR (100 MHz, CD$_3$OD) δ 145.3, 113.7, 65.5, 39.5, 32.0.

The synthetic procedures of *iso*-allylic-fluorophore, *iso*-allylic-fluorophore NHS ester and 3′-O-allyl-dNTP-*iso*-allyl-fluorophore are the same as described previously for allylic-fluorophore, allylic-fluorophore NHS ester, and 3′-O-allyl-dNTP-allyl-fluorophore, respectively.

***iso*-Allyl-Bodipy-FL-510 (15):**

^{1}H NMR (400 MHz, CD$_3$OD) δ 7.42 (s, 1H), 7.00 (d, *J* = 4.0 Hz, 1H), 6.32 (d, *J* = 4.0 Hz, 1H), 6.21 (s, 1H), 5.06 (s, 1H, one of C=C*H*$_2$), 4.87 (s, 1H, one of C=C*H*$_2$), 4.01 (s, 2H, C*H*$_2$OH), 3.33 (t, *J* = 7.5 Hz, 2H), 3.21 (t, *J* = 7.7 Hz, 2H), 2.59 (t, *J* = 7.7 Hz, 2H), 2.51 (s, 3H, one of ArC*H*$_3$), 2.28 (s, 3H), 2.26 (t, *J* = 7.1 Hz, 2H).

***iso*-Allyl-Bodipy-650 (16):**

^{1}H NMR (400 MHz, CD$_3$OD) δ 8.12 (dd, *J* = 0.9, 3.8 Hz, 1H), 7.63 (m, 3H), 7.54 (d, *J* = 6.4 Hz, 2H), 7.35 (s, 1H), 7.18-7.22 (m, 2H), 7.12 (m, 2H), 7.06 (d, *J* = 8.8 Hz, 2H), 6.85 (d, *J* = 4.2 Hz, 1H), 5.06 (s, 1H, one of C=C*H*$_2$), 4.86 (s, 1H, one of C=C*H*$_2$),4.56 (s, 2H), 4.00 (s, 2H, C*H*$_2$OH), 3.28 (m, 4H), 2.23 (t, *J* = 7.1 Hz, 2H), 2.14 (t, *J* = 7.5 Hz, 2H), 1.49-1.62 (m, 4H), 1.25-1.34 (m, 2H).

***iso*-Allyl-R6G (17):**

^{1}H NMR (400 MHz, CD$_3$OD) δ 8.12 (d, *J* = 8.1 Hz, 1H), 8.05 (dd, *J* = 1.8, 8.1 Hz, 1H), 7.66 (d, *J* = 1.6 Hz, 1H), 7.02

(s, 2H), 6.88 (s, 2H), 5.08 (s, 1H, one of C=CH_2), 4.92 (s, 1H, one of C=CH_2), 4.06 (s, 2H, CH_2OH), 3.48-3.56 (m, 6H), 2.40 (t, J = 7.2 Hz, 2H), 2.13 (s, 6H), 1.36 (t, J = 7.2 Hz, 6H).

***iso*-Allyl-ROX (18):**

^{1}H NMR (400 MHz, CD_3OD) δ 8.03 (d, J = 8.1 Hz, 1H), 7.98 (dd, J = 1.6, 8.1 Hz, 1H), 7.60 (d, J = 1.4 Hz, 1H), 6.75 (s, 2H), 5.08 (s, 1H, one of C=CH_2), 4.91 (s, 1H, one of C=CH_2), 4.05 (s, 2H, CH_2OH), 3.45-3.57 (m, 10H), 3.03-3.10 (m, 4H), 2.64-2.73 (m, 4H), 2.38 (t, J = 7.1 Hz, 2H), 2.04-2.15 (m, 4H), 1.89-1.99 (m, 4H).

References

1. Collins, F. S.; Green, E. D.; Guttmacher, A. E.; Guyer, M. S. "A vision for the future of genomics research". *Nature* **2003**, 422, 835-847.

2. Ju, J.; Li, Z.; Edwards, J.; Itagaki, Y. "Massive parallel method for decoding DNA and RNA". U.S. Patent **2003**, 6,664,079.

3. Meng, Q.; Kim, D. H.; Bai, X.; Bi, L.; Turro, N. J.; Ju, J. "Design and synthesis of a photocleavable fluorescent nucleotide 3'-O-allyl-dGTP-PC-Bodipy-FL-510 as a reversible terminator for DNA sequencing by synthesis". *JACS*, submitted

4. Mitra, R. D.; Shendure, J.; Olejnik, J.; Olejnik, E. K.; Church, G. M. "Fluorescent in situ sequencing on polymerase colonies". *Anal. Biochem.* **2003**, 320, 55-65.

5. Pleasants, J. C.; Guo, W.; Rabenstein, D. L. "A comparative study of the kinetics of selenol/diselenide and thiol/disulfide exchange reactions". *J. Am. Chem. Soc.* **1989**, 111, 6553-6558.

6. Huyghues-Despointes, B. M. P.; Nelson, J. W. "Stabilities of disulfide bond intermediates in the folding of apamin". *Biochemistry* **1992**, 31, 1476-1483.

7. Ruparel, H.; Li, Z.; Bai, X.; Kim, D. H.; Turro, N. J.; Ju, J. "Design and synthesis of a 3'-O-allyl

photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis". *Proc. Natl. Acad. Sci. USA* **2005**, 102, 5932-5937.

8. Kang, J.-H.; Siddiqui, M.A.; Sigano, D.M.; Krajewski, K.; Lewin, N.E.; Pu, Y.; Blumberg, P.M.; Lee, J. and Marquez, V.E. Conformationally constrained analogues of diacylglycerol. 24. Asymmetric synthesis of a chiral (R)-DAG-Lactone template as a versatile precursor for highly functionalized DAG-Lactones. *Org. Lett.* **2004,** *6(14)*, 2413-2416

9. Honma, M.; Sawada, T.; Fujisawa, Y.; Utsugi, M.; Watanabe, H.; Umino, A.; Matsumura, T.; Hagihara, T.; Takano, M. and Nakada, M. Asymmetric catalysis on the intramolecular cyclopropanation of α-diazo-β-keto Sulfones *J. Am. Chem. Soc.* **2003,** *125(10)*, 2860-2861.

10. Soli, E.D.; Manoso, A.S.; Patterson, M.C. and DeShong, P. Azide and cyanide displacements via hypervalent silicate intermediates *J. Org. Chem.* **1999,** *64(9)*, 3171-7.

11. Jacobson, M.A. and Williard, P.G. Generation of 1-azapentadienyl anion from *N*-(*tert*-butyldimethylsilyl)-3-buten-1-amine *J. Org. Chem.* **2002,** *67(11)*, 3915-3918.

What is claimed is:

1. A nucleotide analogue comprising (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine, thymine or an analogue of thymine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker.

2. The nucleotide analogue of claim 1, wherein the nucleotide analogue is an analogue of dATP, dGTP, dCTP or dUTP.

3. The nucleotide analogue of claim 1, wherein the fluorophore is selected from the group consisting of ROX, Bodipy-FL-510, Bodipy-650 and R6G.

4. The nucleotide analogue of claim 1, wherein the fluorophore is bound to the base via an *iso*-allyl linker.

5. The nucleotide analogue of claim 4, wherein the nucleotide analogue is selected from the group consisting of 3′-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-*iso*-allyl-Bodipy-650, 3′-O-allyl-dATP-*iso*-allyl-ROX and 3′-O-allyl-dUTP-*iso*-allyl-R6G.

6. The nucleotide analogue of claim 1, wherein the nucleotide analogue is selected from the group consisting of 3′-O-allyl-dGTP-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-allyl-Bodipy-650, 3′-O-allyl-dATP-allyl-ROX and 3′-O-allyl-dUTP-allyl-R6G.

7. A method for making a nucleotide analogue wherein the nucleotide analogue comprises (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′ oxygen of the deoxyribose, and (iv) a fluorophore bound to the base via an allyl linker which is not an *iso*-allyl linker, comprising the steps of:
   (a) contacting 6-amino-hex-2-en-1-ol and an N-hydroxysuccinimide ester of a fluorophore in the presence of a first suitable solvent and a suitable base;
   (b) treating the resulting product of step (a) with DSC/Et_3N in a second suitable solvent; and
   (c) treating the resulting product of step (b) with a 3′-O-allyl-dNTP-NH_2 in the presence of a suitable buffered solvent, wherein the base of the 3′-O-allyl-dNTP-NH_2 is an adenine, guanine, cytosine, uracil, or an analogue thereof, thereby making the nucleotide analogue.

8. The method of claim 7, wherein the steps comprise those set forth in Figure 2, scheme A; Figure 2, scheme B; Figure 2, scheme C; or Figure 2, scheme D.

9. The method of claim 7, wherein the first suitable solvent is DMF and the second suitable solvent is DMF.

10. The method of claim 7, wherein the first suitable solvent is acetonitrile and the second suitable solvent is DMF.

11. The method of claim 7, wherein the suitable base is $NaHCO_3$.

12. The method of claim 7, wherein the suitable buffered solvent is DMF buffered with $NaHCO_3$-Na_2CO_3.

13. The method of claim 7, wherein the suitable buffered solvent is acetonitrile buffered with $NaHCO_3$-Na_2CO_3.

14. A method for making a nucleotide analogue wherein the nucleotide analogue comprises (i) a base selected from the group consisting of adenine or an analogue of adenine, guanine or an analogue of guanine, cytosine or an analogue of cytosine and uracil or an analogue of uracil, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3′ oxygen of the deoxyribose, and (iv) a fluorophore bound to the base via an allyl linker, comprising the steps of:
   (a) contacting 2-(2-amino-ethyl)-prop-2-en-1-ol and an N-hydroxysuccinimide ester of a

fluorophore in the presence of a first suitable solvent and a suitable base;

(b) treating the resulting product of step (a) with DSC/Et_3N in a second suitable solvent; and

(c) treating the resulting product of step (b) with a 3'-O-allyl-dNTP-NH_2 in the presence of a suitable buffered solvent, wherein the base of the 3'-O-allyl-dNTP-NH_2 is an adenine, guanine, cytosine, uracil, or an analogue thereof, thereby making the nucleotide analogue.

15. The method of claim 14, wherein the steps comprise those set forth in Figure 6, scheme A; Figure 6, scheme B; Figure 6, scheme C; or Figure 6, scheme D.

16. The method of claim 14, wherein the first suitable solvent is DMF and the second suitable solvent is DMF.

17. The method of claim 14, wherein the suitable base is $NaHCO_3$.

18. The method of claim 14, wherein the suitable buffered solvent is DMF buffered with $NaHCO_3$-Na_2CO_3.

19. A method for determining the sequence of a DNA comprising performing the following steps for each residue of the DNA to be sequenced:

(a) contacting the DNA with a DNA polymerase in the presence of (i) a primer and (ii) four

fluorescent nucleotide analogues under conditions permitting the DNA polymerase to catalyze DNA synthesis, wherein (1) the nucleotide analogues consist of an analogue of dGTP, an analogue of dCTP, an analogue of dTTP or dUTP, and an analogue of dATP, (2) each nucleotide analogue comprises (i) a base selected from the group consisting of adenine, guanine, cytosine, thymine or uracil, and analogues thereof, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3'-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker, so that a nucleotide analogue complementary to the residue being sequenced is bound to the DNA by the DNA polymerase, and (3) each of the four analogues has a predetermined fluorescence wavelength which is different than the fluorescence wavelengths of the other three analogues;

(b) removing unbound nucleotide analogues;

(c) determining the identity of the bound nucleotide analogue; and

(d) following step (c), except with respect to the final DNA residue to be sequenced, chemically cleaving from the bound nucleotide analogue the fluorophore and the allyl moiety bound to the 3'-oxygen atom of the deoxyribose,

thereby determining the sequence of the DNA.

20. The method of claim 19, wherein chemically cleaving the fluorophore and the allyl moiety bound to the 3′-oxygen atom is performed using Na_2PdCl_4.

21. The method of claim 19, wherein the primer is a self-priming moiety.

22. The method of claim 19, wherein the DNA is bound to a solid substrate.

23. The method of claim 22, wherein the DNA is bound to the solid substrate via 1,3-dipolar azide-alkyne cycloaddition chemistry.

24. The method of claim 22, wherein about 1000 or fewer copies of the DNA are bound to the solid substrate.

25. The method of claim 19, wherein the four fluorescent nucleotide analogues are 3′-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-*iso*-allyl-Bodipy-650, 3′-O-allyl-dATP-*iso*-allyl-ROX and 3′-O-allyl-dUTP-*iso*-allyl-R6G.

26. The method of claim 19, wherein the four fluorescent nucleotide analogues are 3′-O-allyl-dGTP-allyl-Bodipy-FL-510, 3′-O-allyl-dCTP-allyl-Bodipy-650, 3′-O-allyl-dATP-allyl-ROX and 3′-O-allyl-dUTP-allyl-R6G.

27. The method of claim 19, wherein the DNA polymerase is a 9°N polymerase.

28. A kit for performing the method claim 19, comprising, in separate compartments,
   (a) a nucleotide analogue of (i) GTP, (ii) ATP, (iii) CTP and (iv) TTP or UTP, wherein each analogue comprises (i) a base selected from the group consisting of adenine, guanine, cytosine, thymine or uracil, or an analogue thereof, (ii) a deoxyribose, (iii) an allyl moiety bound to the 3'-oxygen of the deoxyribose and (iv) a fluorophore bound to the base via an allyl linker,
   (b) reagents suitable for use in DNA polymerization; and
   (c) instructions for use.

29. The kit of claim 28, wherein the kit comprises 3'-O-allyl-dGTP-allyl-Bodipy-FL-510, 3'-O-allyl-dCTP-allyl-Bodipy-650, 3'-O-allyl-dATP-allyl-ROX and 3'-O-allyl-dUTP-allyl-R6G.

30. The kit of claim 28, wherein the kit comprises 3'-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510, 3'-O-allyl-dCTP-*iso*-allyl-Bodipy-650, 3'-O-allyl-dATP-*iso*-allyl-ROX and 3'-O-allyl-dUTP-*iso*-allyl-R6G.

31. A method for covalently affixing a detectable moiety, via an allyl linker, to an NH_2-bearing molecule, comprising contacting the detectable moiety with the the NH_2-bearing molecule in the presence of a suitable solvent and suitable base, wherein the detectable moiety comprises a mass tag,

fluorophore or chromophore bound to a NHS ester of an allyl moiety.

32. The method of claim 31, wherein the NH_2-bearing molecule is a nucleotide and the detectable moiety comprises a fluorophore.

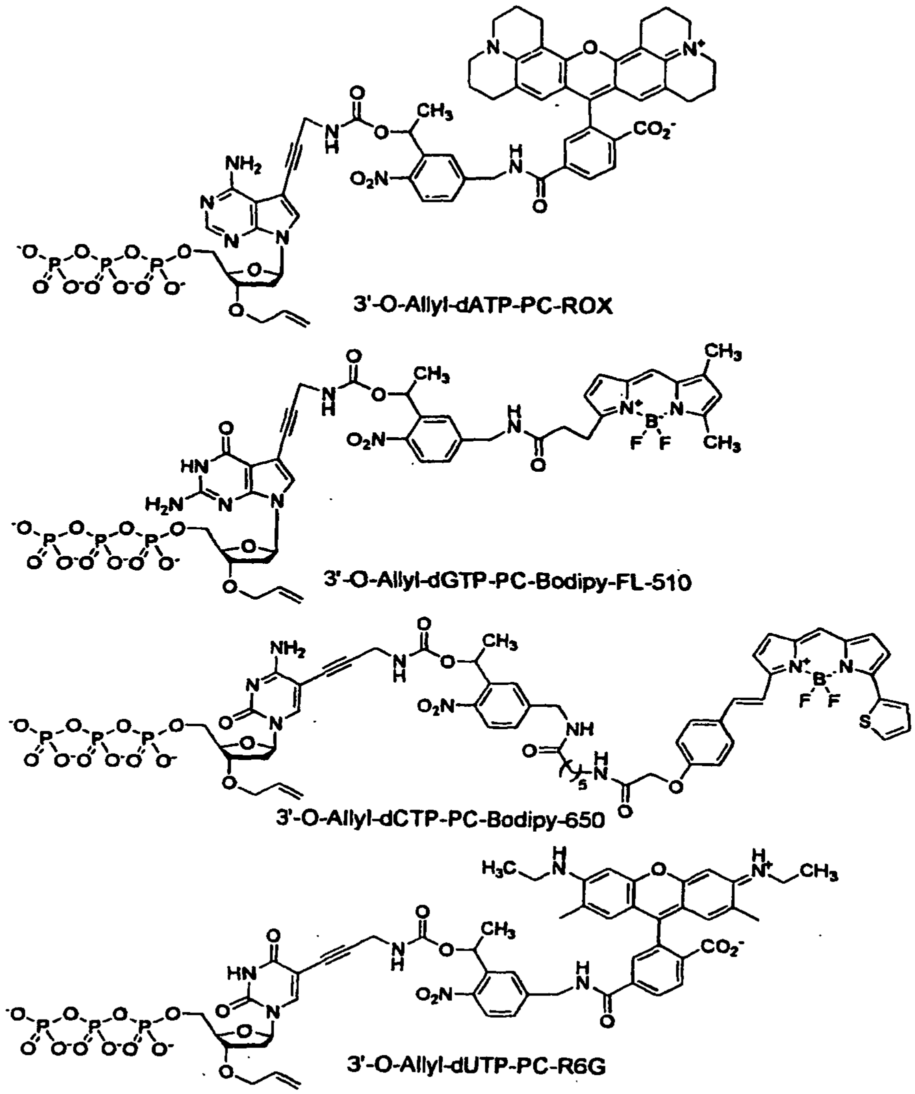
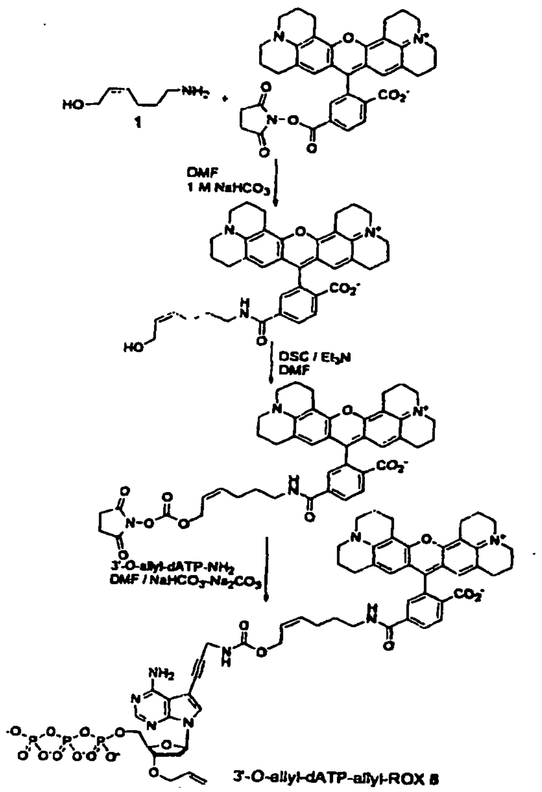
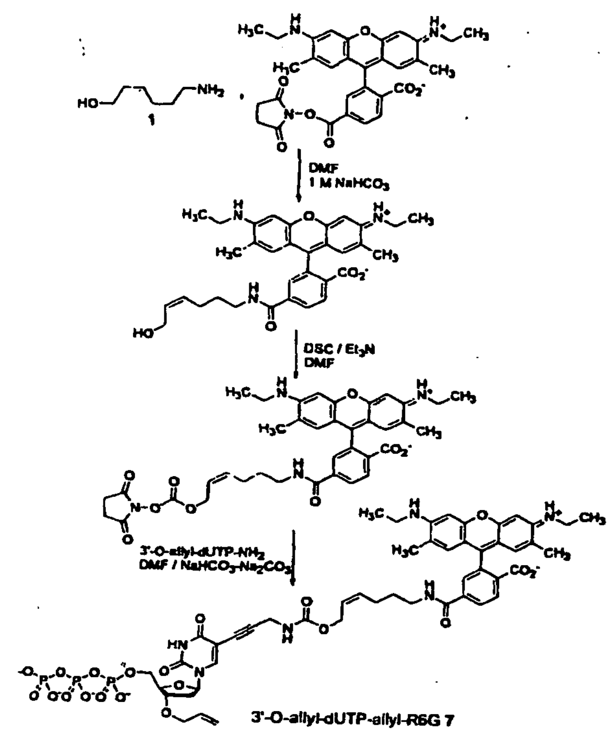

3'-O-Allyl-dATP-PC-ROX
3'-O-Allyl-dGTP-PC-Bodipy-FL-510
3'-O-Allyl-dCTP-PC-Bodipy-650
3'-O-Allyl-dUTP-PC-R6G

Fig. 1

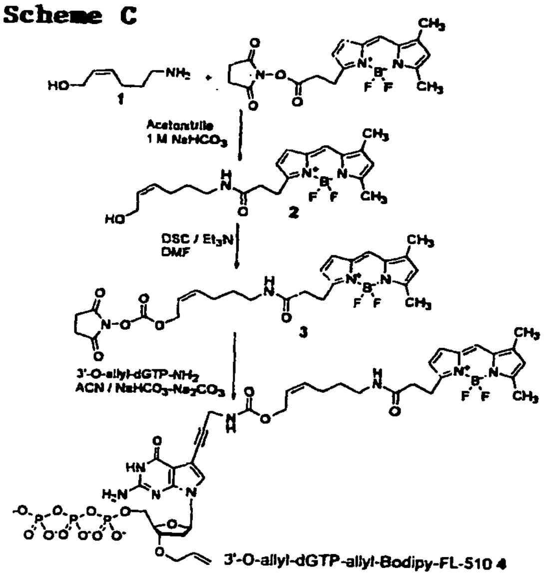
Scheme A



Scheme B



Scheme C



Scheme D

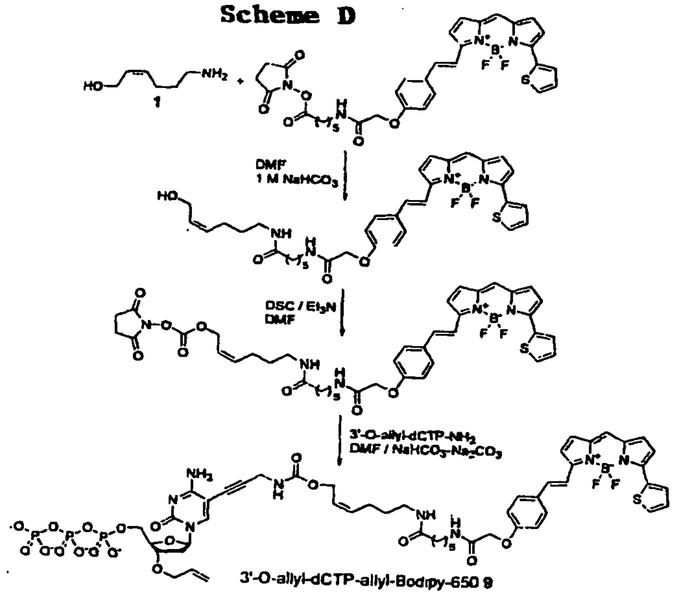


Fig. 2

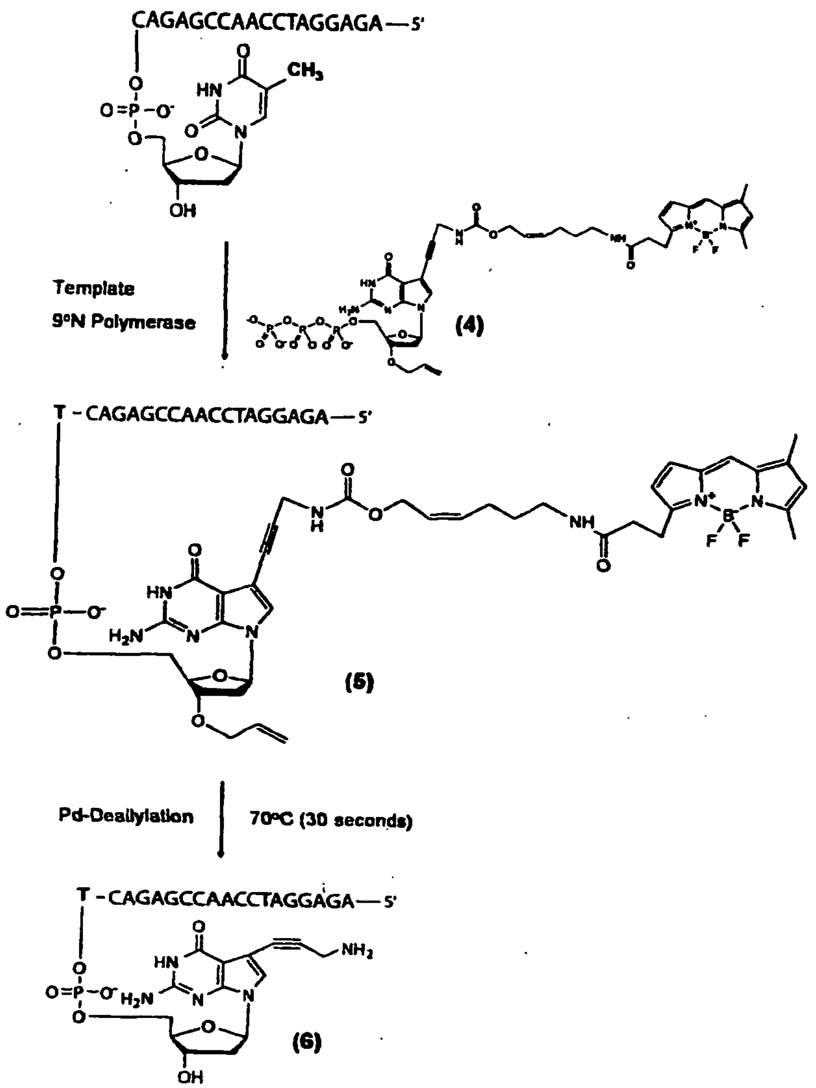

CAGAGCCAACCTAGGAGA—5'
Template
9°N Polymerase
(4)
T - CAGAGCCAACCTAGGAGA—5'
(5)
Pd-Deallylation
70°C (30 seconds)
T - CAGAGCCAACCTAGGAGA—5'
(6)

Fig. 3

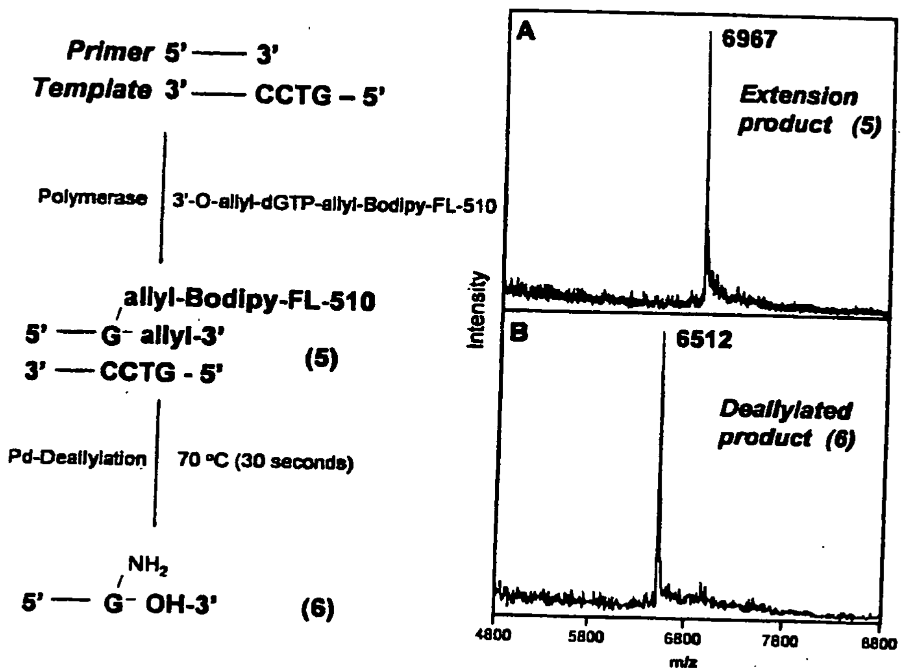
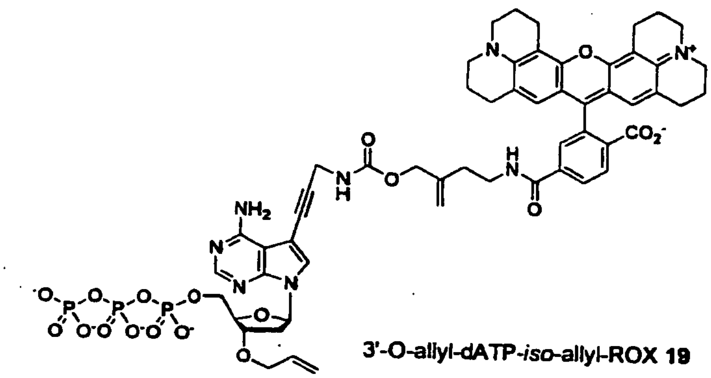
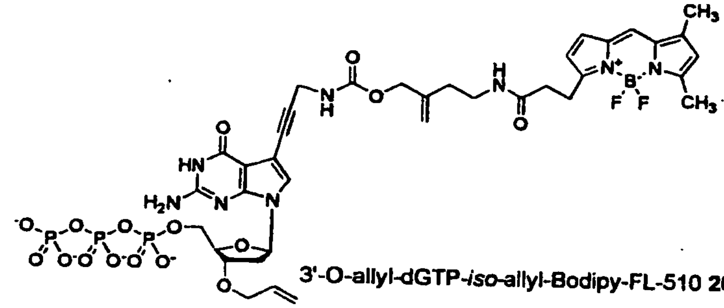

Primer 5' — 3'
Template 3' — CCTG – 5'
Polymerase
3'-O-allyl-dGTP-allyl-Bodipy-FL-510
allyl-Bodipy-FL-510
5' — G- allyl-3'
3' — CCTG - 5'
(5)
Pd-Deallylation
70 °C (30 seconds)
NH2
5' — G- OH-3'
(6)
A
6967
Extension product (5)
B
6512
Deallylated product (6)
Intensity
4800
5800
6800
7800
8800
m/z

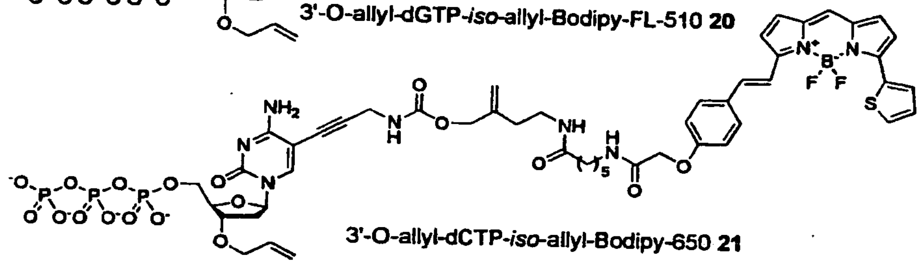
Fig. 4



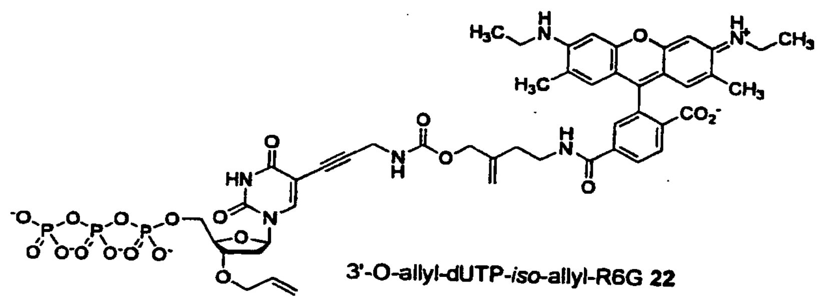
3'-O-allyl-dATP-*iso*-allyl-ROX **19**



3'-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510 **20**



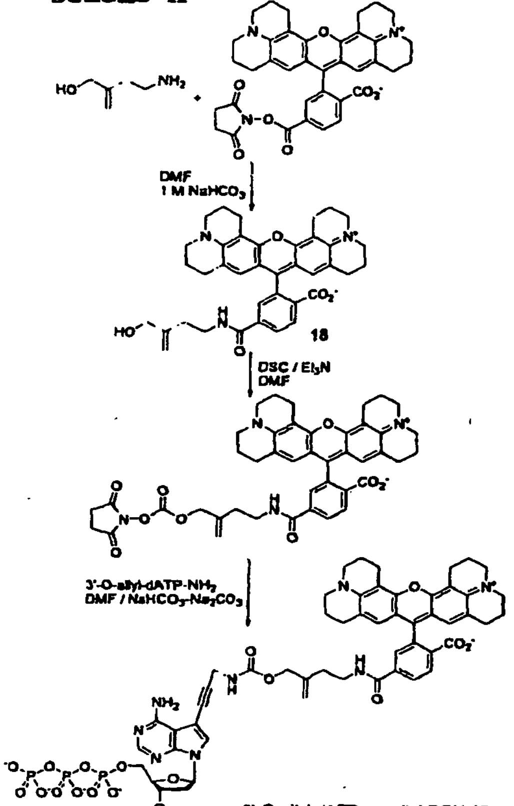
3'-O-allyl-dCTP-*iso*-allyl-Bodipy-650 **21**



3'-O-allyl-dUTP-*iso*-allyl-R6G **22**

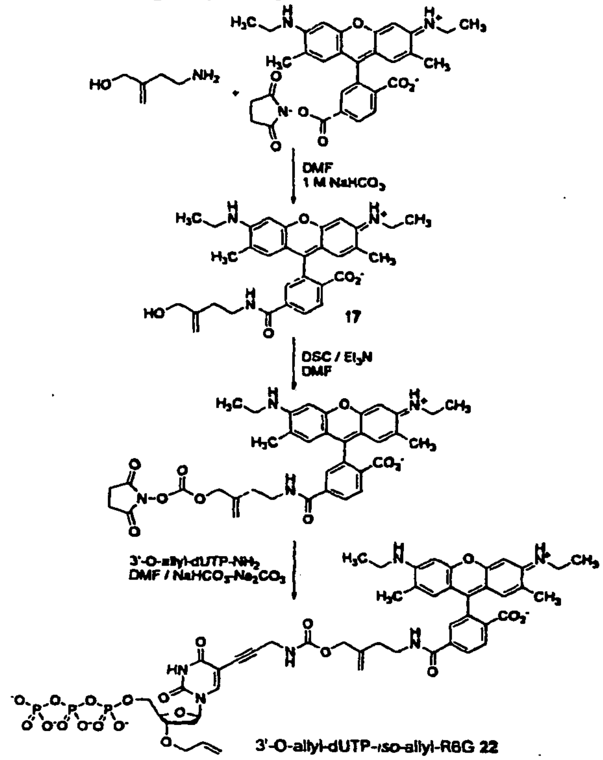
Fig. 5

Scheme A



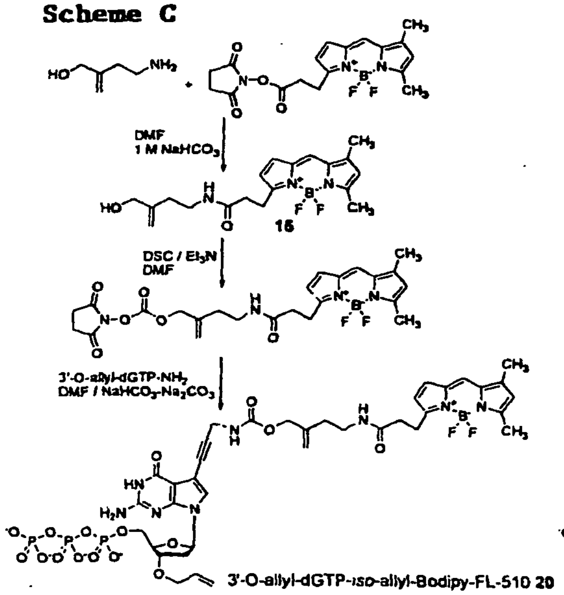
3'-O-allyl-dATP-*iso*-allyl-ROX **19**

Scheme B



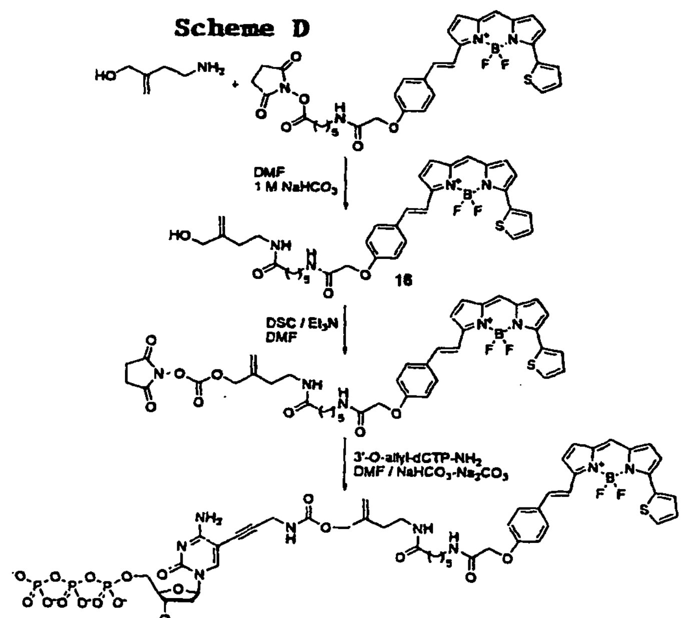
3'-O-allyl-dUTP-*iso*-allyl-R6G **22**

Scheme C



3'-O-allyl-dGTP-*iso*-allyl-Bodipy-FL-510 **20**

Scheme D



3'-O-allyl-dCTP-*iso*-allyl-Bodipy-650 **21**

Fig. 6

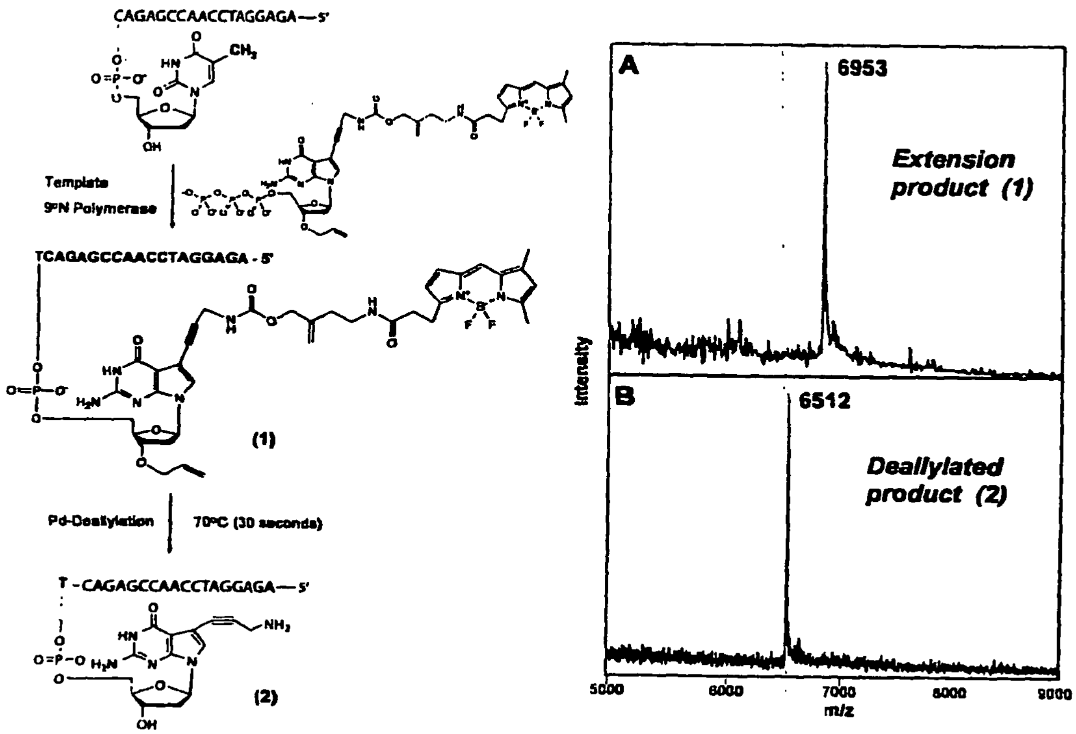

CAGAGCCAACCTAGGAGA—5'
Template
9°N Polymerase
TCAGAGCCAACCTAGGAGA - 5'
(1)
Pd-Deallylation
70°C (30 seconds)
T - CAGAGCCAACCTAGGAGA—5'
(2)
A
6953
Extension product (1)
B
6512
Deallylated product (2)
Intensity
5000
6000
7000
8000
9000
m/z

Fig. 7

SEQUENCE LISTING

<110> The Trustees of Columbia University in the City of New York
Ju, Jingyue
Bi, Lanrong
Dae, Kim H.
Meng, Qinling

<120> CHEMICALLY CLEAVABLE 3'-O-ALLYL-DNTP-ALLYL-FLUOROPHORE FLUORESCENT NUCLEOTIDE ANALOGUES AND RELATED METHODS

<130> 0575/71545-A-PCT/JPW/YC

<150> US 60/732,040
<151> 2005-10-31

<160> 2

<170> PatentIn version 3.3

<210> 1
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer directed to Human p53

<400> 1
gttgatgtac acattgtcaa 20

<210> 2
<211> 100
<212> DNA
<213> Homo sapiens

<400> 2
tacccggagg ccaagtacgg cgggtacgtc cttgacaatg tgtacatcaa catcacctac 60

caccatgtca gtctcggttg gatcctctat tgtgtccggg 100